



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Children's sensitivity to phonological and semantic cues during noun class learning

Citation for published version:

Culbertson, J, Jarvinen, H, Haggarty, F & Smith, K 2019, 'Children's sensitivity to phonological and semantic cues during noun class learning: Evidence for a phonological bias', *Language*, vol. 95, no. 2, pp. 268-293.
<https://doi.org/10.1353/lan.0.0234>

Digital Object Identifier (DOI):

[10.1353/lan.0.0234](https://doi.org/10.1353/lan.0.0234)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Language

Publisher Rights Statement:

This is an Accepted Manuscript of an article published in "Language" on 21.05.2019 available online:
<https://doi.org/10.1353/lan.0.0234>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





PROJECT MUSE®

Children's Sensitivity to Phonological and Semantic Cues
During Noun Class Learning: Evidence for a Phonological Bias

Jennifer Culbertson, Hanna Jarvinen, Frances Haggarty, Kenny Smith

Language, Ahead of Print, vol. 95, no. 2, (Article)

Published by Linguistic Society of America

DOI: <https://doi.org/10.1353/lan.0.0234>



This is a preprint article. When the final version of this article launches, this URL will be automatically redirected.

➔ For additional information about this preprint article

<https://muse.jhu.edu/article/725764/summary>

CHILDREN’S SENSITIVITY TO PHONOLOGICAL AND SEMANTIC CUES
DURING NOUN CLASS LEARNING: EVIDENCE FOR A PHONOLOGICAL BIAS

JENNIFER CULBERTSON

University of Edinburgh

HANNA JARVINEN

University of Edinburgh

FRANCES HAGGARTY

University of Edinburgh

KENNY SMITH

University of Edinburgh

Previous research on the acquisition of noun classification systems (e.g. grammatical gender) has found that child learners rely disproportionately on phonological cues to determine the class of a new noun, even when competing semantic cues are more reliable in their language. Culbertson, Gagliardi, and Smith (2017) use artificial language learning experiments with adults to argue that this likely results from the early availability of phonological information during acquisition. Learners base their initial representations on formal features of nouns, only later integrating semantic cues from noun meanings. Here, we use these same methods to show that early availability affects cue use in children (six- to seven-year-olds) as well. However, we also find evidence of developmental changes in sensitivity to semantics; when both cue types are simultaneously available, children are more likely to rely on phonology than adults are. Our results suggest that both early availability and a bias favoring phonological cues contribute to children’s overreliance on phonology in natural language acquisition.*

Keywords: language acquisition, gender, noun classification, artificial language learning, phonology, semantics

1. INTRODUCTION. Noun classification systems, such as grammatical gender, are widespread across the world’s languages, and they have been extensively studied from the perspective of both linguistic typology and language acquisition. Examples from three languages are shown in 1, 2, and 3 below. The first, French, has a gender¹ system in which nouns fall into two classes, MASCULINE and FEMININE; the form of the article (as well as other elements in the noun phrase) varies based on noun gender, for example, *le* for masculine nouns, *la* for feminine nouns. The second example, Tsez (a Nakh-Dagestan language spoken in the Caucasus), has four noun classes and exhibits agreement within the noun phrase as well as class-based verbal agreement. The third is Cantonese, which has a numeral classifier system consisting of many distinct forms, used with nouns depending on semantic features (the classifier *go*[3] is used for humans, *zek*[3] is used primarily for animals, *zi*[1] is used for long, narrow objects, and *jeung*[1] is used for flat objects; Tsang & Chambers 2011).

(1) French

- | | |
|--|--|
| a. <i>le</i> <i>garçon</i>
the.M boy | b. <i>le</i> <i>stylo blanc</i>
the.M pen white.M |
| c. <i>la</i> <i>filles</i>
the.F girl | d. <i>la</i> <i>tasse blanche</i>
the.F cup white.F |

* This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement Nos. 681942 and 757643). We would like to thank all of the parents and children who participated in these studies.

¹ Linguists and typologists traditionally distinguish these three types of classification systems: gender, noun class, and noun (or numeral) classifiers. The major formal distinction between gender/noun class and classifiers is that the former trigger agreement while the latter do not (Corbett 1991). Here we focus mainly on noun class systems and use this term to include systems traditionally called ‘gender’.

(2) Tsez

- a. igu baba
 good.1 father
 c. bigu buq
 good.3 sun

- b. jigu kid
 good.2 girl
 d. rigu čorpa
 good.4 soup

(3) Cantonese

- a. sam[1] go[3] jan[4]
 three CL people
 c. sam[1] zi[1] bat[1]
 three CL pens

- b. sam[1] zek[3] gau[2]
 three CL dogs
 d. sam[1] jeung[1] zi[2]
 three CL paper

Work on the typology of noun classification systems has documented extensive variation in the formal characteristics of such systems. However, a number of recurrent features have also been documented. For example, across different types of noun classification systems—including grammatical gender, noun class, and numeral classifier systems—semantic features of noun referents regularly play a role in determining how nouns are classified. Indeed, the same semantic features appear in a great many, if not all, noun classification systems: natural gender, animacy, and shape (Denny 1976, Dixon 1986, Lakoff 1987, Comrie 1989, Aikhenvald 2000, Senft 2000). This can be seen in the three examples above: natural gender is a cue to class in both French and Tsez, animacy is a cue to class in both Tsez and Cantonese, and shape is a cue to class in Cantonese. In addition to semantic cues, phonological features of the noun itself may also be used to determine a noun's class. This is again illustrated in our examples: French nouns ending in *-on* are more likely to be masculine, while those ending in *-asse* are more likely to be feminine; in Tsez both noun-initial (*b-*, *r-*) and noun-final (*-i*) segments provide a cue to class: for example, nouns beginning in *b-* are more likely to be in class 3. However, while semantic cues are always present, phonological cues are found only in a subset of noun classification systems and are not used at all in numeral classifier systems like Cantonese (Corbett 1991, Aikhenvald 2000).²

Crosslinguistically, French and Tsez represent typical noun class systems, which use a mix of semantic and phonological class cues. Such cues are sometimes highly reliable, and sometimes merely indicate a higher probability of one class or another. In French, natural gender, when it is available, provides a near-deterministic cue to grammatical gender. However, looking at nouns with no natural gender reveals a host of additional nondeterministic (and therefore less reliable) cues, many of them phonological: for example, certain endings of nouns (e.g. *-elle*) provide a strong cue to feminine gender, while others (e.g. *-ire*) are much weaker (Lyster 2006, Ayoun 2018).³ Similarly, natural gender and animacy are highly reliable predictors of class in Tsez (all and only human males are class 1, all human females are class 2, and so on), while phonological features of the nouns themselves are less reliable. Indeed, computational models have been used to confirm that semantic cues to class in Tsez are more predictive than phonological cues (Plaster, Polinsky, & Harizanov 2013, Gagliardi & Lidz 2014).

To summarize, work on the typology of noun classification systems suggests that semantic cues to class are ubiquitous; all such systems make use of semantic cues, they are drawn from a recurrent set, and they are often highly reliable. By contrast, phonological

² In the *World atlas of language structures*, Corbett (2013) lists fifty-three languages with semantic-based gender assignment, and fifty-nine with a mix of semantic and phonological cues.

³ See e.g. <https://www.fourmilab.ch/francais/gender.html>, <https://french.stackexchange.com/questions/2616/is-there-any-general-rule-to-determine-the-gender-of-a-noun-based-on-its-spellin>.

cues are only sometimes present and appear to have a wider range of reliability; furthermore, and unlike for semantic cues, there are no obvious crosslinguistic patterns in terms of the types of phonological features used (though noun beginnings and endings are a common position in which such cues occur). This overall asymmetry between semantics and phonology is particularly interesting from the perspective of acquisition. Noun class systems present a significant challenge to both first and second language learners (MacWhinney 1978, Braine 1987, Levy 1988, Frigo & McDonald 1998, Carroll 1999, Kempe & Brooks 2001, Taraban 2004, Arnon & Ramscar 2012, among others). They require learners to track both deterministic and probabilistic cues across multiple domains, each of which may apply to a different subset of nouns, with a substantial number of exceptional elements that must simply be memorized. On the face of it, semantic cues like natural gender, animacy, and shape are highly salient and early acquired (Bunger & Lidz 2006, Becker 2009, Strickland 2017). Therefore, if these types of cues are present for a subset of nouns in a language and are highly reliable indicators of a noun's class, then such cues should be extremely useful to young learners. However, previous research on natural language acquisition of noun classification systems has shown that child learners in fact appear to rely disproportionately on phonological cues.

Karmiloff-Smith (1981) found that when French children were faced with conflicting semantic and phonological cues in a novel noun, they used the phonological information to determine grammatical gender. She taught French children novel nouns denoting female and male alien characters. In some cases, the phonological form of the label was designed to conflict with the cue from natural gender: a noun ending in *-elle* (typically found on feminine nouns) was used for a male alien, or a noun ending in *-on* (typically found on masculine nouns) was used for a female alien. While adult intuitions indicated that natural gender should override phonology, children as young as three and up to age ten were found to use the definite article that matched the phonological features of the noun, rather than natural gender. This same finding has since been documented across a number of languages (Czech: Henzl 1975; German: Mills 1985; Hebrew: Levy 1983; Russian: Rodina & Westergaard 2012; Sesotho (Bantu): Demuth 2000, Demuth & Ellis 2008; Spanish: Perez-Péreira 1991, Mariscal 2009), including Tsez, where the relevant semantic cues have explicitly been shown to be more reliable predictors of class (including in child-directed speech; Gagliardi & Lidz 2014); young Tsez children nevertheless appear to overrely on phonology, while Tsez adults do not.

1.1. EXPLAINING AN APPARENT BIAS FOR PHONOLOGY. Why might child learners exhibit an overreliance on phonological cues, even when such cues are sometimes less reliable? A number of potential explanations for this have been proposed in the literature. First, it could be that in these languages the semantic cues, although reliable, are not very perceptually salient. As mentioned above, however, the semantic cues in question tend to reflect early-acquired features like animacy and natural gender, while the phonological cues are often low-salience word endings (Frigo & McDonald 1998; see also Culbertson et al. 2017, described below). Second, it could be that although semantic cues are not systematically less reliable or perceptually salient, they are restricted to a small subset of items in the language and thus are not available to act as a cue with high frequency. For example, the vast majority of nouns do not have a natural gender, so although this cue is highly predictive of class in French, it may not be as useful. However, the same overreliance on phonology is found in Tsez, where the combination of natural gender and animacy covers a large portion of the noun space. Moreover, at least in French and Tsez, particular phonological cues are not obviously more frequent than their semantic counterparts. A third possible explanation is that phonological cues

are available (or accessible) to the learner before semantic cues; surface phonological features of dependent elements like determiners and nouns can in principle be learned well before word meanings are acquired (Carroll 1999, Polinsky & Jackson 1999, Demuth 2000, Culbertson & Wilson 2013, Gagliardi et al. 2017). Finally, it could be that children's overreliance on phonology is due to an active bias against using external cues like semantics in forming grammatical categories, particularly when internal, phonological, properties of nouns are available to cue class (Gagliardi 2012, Culbertson & Wilson 2013, Gagliardi & Lidz 2014, Gagliardi et al. 2017). This view predicts that learners have an *a priori* bias in favor of phonological cues when they are present.

Culbertson, Gagliardi, and Smith (2017) argue that overreliance on phonology is likely due to its early availability to child learners rather than to an *a priori* bias favoring phonology over semantics. Their claim is that if the noun classification system is initially built on the basis of these early-available phonological cues, semantic cues acquired later might take time to be integrated into the system. To test this claim, they modeled the availability of different types of cues in a series of artificial language learning experiments with adults. The miniature artificial noun class system featured two classes, indicated by two distinct class markers (independent words, akin to determiners). Class membership was deterministically cued by both a phonological cue on the noun and a semantic cue from the noun referent. The two cues were deliberately confounded, so that in principle both provided an equally reliable cue to class (Figure 1A).⁴ At test (Figure 1B), learners were given new nouns in which the previously confounded cues conflicted—for example, a noun with the phonological feature of one class, but the semantic feature of the other class.

In the first set of experiments, both cues were available simultaneously. Adults in these experiments showed no *a priori* preference for using phonology over semantic cues when choosing which class a novel noun belonged to. Rather, perhaps unsurprisingly, adults' reliance on a given cue depended on its perceptual salience. For example, when the semantic cue was animacy and the phonological cue was a suffix on the noun, adults were more likely to choose class markers based on the semantic cue (Figure 1D); in contrast, when the semantic cue was more subtle, involving a distinction based on flexibility (e.g. 'rope', 'paper' in class 1, 'pole', 'tray' in class 2), and the phonological cue involved both a prefix and a suffix, adults based their classification decisions on the more salient phonological cue.

To test the role of cue availability, Culbertson, Gagliardi, and Smith (2017) then conducted subsequent experiments in which exposure to semantic and phonological cues was STAGED; during an initial learning phase, only one cue was available (e.g. Figure 1C). They found that when either the phonological OR the semantic cue was available first, adult learners prioritize the earlier-learned cue, even if it was *a priori* less salient than the competing but late-available cue. Most relevantly, if the initial learning phase included only the word forms and no evidence for word meaning (e.g. phonological cues are available before semantic cues, which they argue is the default ordering of cue availability), then even after significant later exposure to the highly salient semantic cue, adults still chose class markers based on less-salient phonological cues (Fig. 1D). While these results are based on adult learning, they are consistent with the mechanism proposed above for children's overreliance on phonology: children start building their

⁴ Note that the term 'confounded' here refers to situations when one cannot determine the independent effect of one variable because it always cooccurs with the other. This exactly captures the problem that the training items are intended to pose for a learner who wants to know which cue predicts noun class.

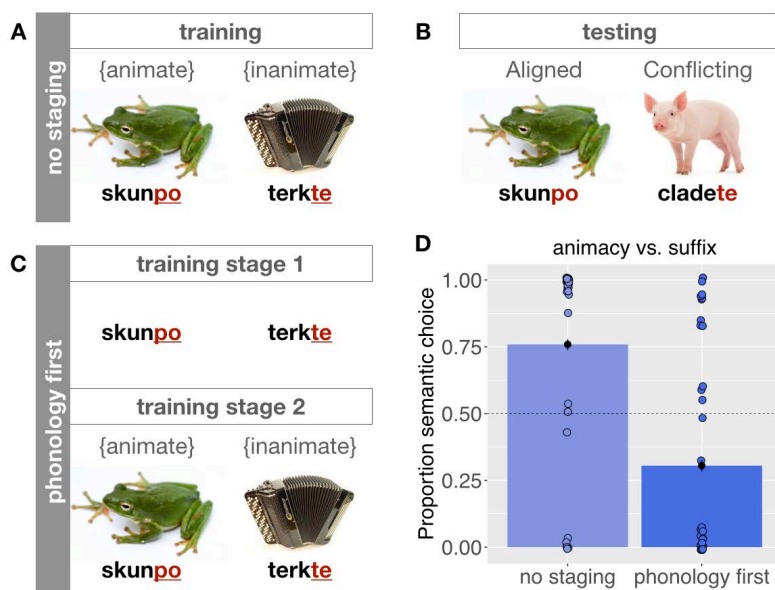


FIGURE 1. Summary of key results from Culbertson et al. 2017. A: Example training trials for condition with animacy and suffix as semantic and phonological cues, respectively. B: Example testing trials, with aligned trials, which conformed to the training stimuli on both cues, and conflicting trials, in which the two cues conflicted. C: Example trials from a staged condition, here phonology-first, showing the two training stages. In the first, only the labels (along with the phonological cue) appeared. The second stage and the testing phase were identical to the unstaged condition. D: Plot showing results of the no-staging and phonology-first conditions for these two cues, showing how often participants chose a class marker that conformed to the semantic cue they were trained on. When both cues were available from the start ('no staging' results), participants preferred to base their classification on the more salient semantic cue; however, when the phonological cue (the relatively less salient suffix) was available first, participants were more likely to base their classification on that phonological cue.

classifications systems very early, when phonological information is available but word meanings are not, and therefore they are more reliant on phonological cues. In other words, these results suggested that children's preference for phonological cues does not necessarily reflect an a priori preference for phonology but could instead be explained as a consequence of cue early-availability.

Here we use a similar set of experiments to investigate whether these results also hold for child learners. That is, we explore whether children have an a priori bias for phonological or semantic cues, and whether they are sensitive to cue early-availability in the same way that the adults in Culbertson et al. 2017 were. In a series of experiments, we show that two cues—one semantic and one phonological—that child learners are equally sensitive to in isolation are, in fact, treated differently when they are in conflict. In particular, unlike adults, children in our experiments show evidence of prioritizing phonological cues even when semantic cues are equally reliable and in principle equally available. Nevertheless, staging cue availability can amplify children's use of both types of cues, just as it does for adults. This suggests, contrary to the argument developed in Culbertson et al. 2017 (and against our own expectations), that children are in fact biased to attend to phonological cues when acquiring noun classes. An a priori bias for phonology and earlier availability of phonological cues may therefore in combination lead to the robust finding of children's overreliance on phonology at early stages of acquisition of noun classification systems across natural languages.

The article proceeds as follows: in experiment 1 we first establish a set of phonological and semantic cues that children (and adults) can learn well in isolation. In experiment 2 we then introduce a confound between the cues in training in order to test which cue is relied upon when they conflict at test (as in Culbertson et al. 2017). Finally, in experiment 3 we manipulate the staging of each cue in order to test for the effects of cue early-availability.

2. EXPERIMENT 1. Our ultimate aim is to test whether children prioritize phonological or semantic cues when both are available, and whether children are sensitive to cue early-availability. However, we first need to identify a set of specific cues that children can in principle use to learn an artificial noun class system in the lab. Previous research suggests that children are able to rapidly learn semantically cued (animacy-based) artificial noun class systems with high accuracy when those semantic cues are deterministic (Brown et al. 2018);⁵ in experiment 1 we use a similar semantic cue and test for an equivalent novel phonological cue. We taught children a miniature artificial noun class system in which class was deterministically cued by either semantic information about noun referents OR phonological information from the form of the noun itself. External evidence for the two classes in the language came from a plural marker that differed by class. The semantic cue we used was animacy-based, following Brown et al. (2018) and other previous work showing that animacy is a highly salient cue for adult learners in the context of artificial noun class learning (Culbertson & Wilson 2013, Culbertson et al. 2017). The phonological cue we used was a unique vowel sound that occurred in all of the nouns in each class. These vowels were reduplicated in the stems: for one class, the vowel ‘a’ appeared twice in each stem (e.g. *pata*), and for the other class, the vowel ‘i’ appeared twice in each stem (e.g. *mipi*). Since prior work has shown that reduplication aids word learning in children (Ota & Skarabela 2016, 2017),⁶ this cue is also likely to be highly salient to children. Since these exact cues have not previously been tested in adults, and because the method we are using here differs to some degree from those used in Culbertson et al. 2017 (see below), we also tested adults using the same paradigm that we used for children.

2.1. METHOD. The method we use is an artificial language learning version of the classic ‘wug test’ (Berko 1958), using a procedure closely modeled on that used by

⁵ By ‘deterministic’ we mean that all nouns of a given class are members of a single semantic category. Brown et al. (2018) show failure to generalize in five- to six-year-olds when exception items from one noun class have semantic features characteristic of the other noun class (e.g. animate nouns generally appear with class marker 1, but a single exceptional animate noun appears with class marker 2); even after four training sessions, children performed no better than chance on novel nouns when such exceptions existed, while adults were able to correctly generalize under such conditions. One possible interpretation of this finding is that exception items that possess a relevant semantic feature but appear in the wrong class are particularly problematic for children. However, Schwab et al. (2018) show a similar failure to generalize when exception items lack the class-typical cue—that is, when three of four nouns in noun class 1 have male human referents, three of four referents in noun class 2 have female human referents, and the exception items in each class are inanimate and lack natural gender. Interestingly, no parallel studies have been conducted with phonological cues; we know of no studies that would allow us to compare children’s acquisition of deterministic and non-deterministic phonological cues to class (for adults see Frigo & McDonald 1998).

⁶ This cue is similar in terms of phonological substance to high-salience phonological cues used in Culbertson et al. 2017 and Frigo & McDonald 1998, which featured, for example, both a prefix and a suffix added to nouns. Here we used reduplication for two reasons: first, we wanted to keep word length to a minimum; second, previous studies have suggested that children have trouble learning classification systems in the lab, so we wanted to create cues that were as high-salience as possible, and there is reason to believe that reduplication might be such a cue for children (Ota & Skarabela 2016, 2017).

Schuler, Yang, and Newport (2016). Children were trained on the singular and plural forms of novel words referring to a set of unfamiliar objects. At test, they were given the singular form of a noun and had to produce the plural. Each class of nouns had a distinct plural marker, which distributionally indicated noun class in the language. Note that, unlike in Culbertson et al. 2017, the class markers are therefore semantically meaningful, indicating plurality; however, in common with these previous studies, learners are not required to learn the noun stems themselves, only the class markers and the mapping between markers and stems.

PARTICIPANTS. Child participants were thirty-nine native English-speaking six- to seven-year-olds (seventeen female; mean age = 6;11, range = 6;1–7;9), who were pupils at two Aberdeen City Council primary schools. Adult participants were forty native English speakers (twenty-seven female), who were students at the University of Edinburgh and who received £5 for their participation. Participants were randomly assigned to one of the two conditions (twenty adults in each condition; twenty children in phonology-only, nineteen in semantics-only⁷). Eleven children were bilingual (two in Russian, and one each in Arabic, Gaelic, German, Nepali, Polish, Spanish, Tamil, Urdu, and Yoruba), and fifteen adults were bilingual (three in German, three in Mandarin, two in French, two in Hindi/Urdu, and one each in Arabic, Czech, Hebrew, Serbian, and Spanish).⁸

STIMULI. The artificial lexicon was composed of twelve C_1VC_2V words, six with the vowel 'i' and six with 'a', with C_2 matching across words with a given vowel (drawn pseudo-randomly from a possible pool of 152 such words; see Figure 2). These words were used to label twelve simple objects, either aliens or planets (see Figure 3). Words were divided into two classes of equal size, as indicated by two distinct plural markers that followed the noun.⁹ Cues to noun class differed across conditions. In the phonology-only condition, the cue to class came from the phonological form of the noun alone (either C_1iC_2i or C_1aC_2a), with the semantic types perfectly balanced (i.e. half aliens, half planets in each class). In the semantics-only condition, the cue to class came from the semantic category of the noun (either alien or planet), with the phonological types perfectly balanced (i.e. half C_1iC_2i , half C_1aC_2a in each class). Words were recorded by a female native speaker of Standard Southern British English.

⁷ No children were excluded from the analysis (all completed the experiment). Two adult participants were excluded: one failed to repeat the words in the training phase, and one used markers that were ambiguous and did not match the markers used in training. Note that, as described below, the experimenter was not sitting in the booth with the adult participants.

⁸ Here we use 'bilingual' to mean having native or substantial fluency in another language. Most of these languages have noun class systems (Nepali has an 'attenuated' gender system, with gender agreement applying only to female animates (Masica 1993); Yoruba has no gender system). Of these, the majority feature a combination of phonological and semantic cues to class (Arabic, Czech, French, Gaelic, German, Hebrew, Hindi/Urdu, Polish, Russian, Serbian, and Spanish). However, the phonological cues in Polish, Spanish, and Urdu apply very widely (e.g. in Spanish, final vowels *-a* and *-o* are present on many nouns and reliably cue class). By contrast, Tamil has purely semantic assignment of gender (Asher 1985), and Mandarin (like Cantonese) has numeral classifiers for which the only cues are semantic in nature. We do not analyze further the effect of prior experience with gender systems, since these systems vary widely and are similarly distributed across our child and adult participants.

⁹ For each participant (in this and all other experiments reported here), the two plural markers were chosen randomly from the following set of possible pairs: [gæ, fʌ], [jæ, gʌ], [zæ, gʌ], [kʌ, poʊ], [gʌ, poʊ], [fʌ, koʊ], [jʌ, poʊ], [gʌ, zoʊ], [kʌ, zoʊ], [gæ, poʊ], [gæ, zoʊ], [jæ, koʊ], [jæ, poʊ], [zæ, koʊ], [zæ, poʊ], where [æ] is the vowel in *hat*, [ʌ] is the vowel in *cut*, [oʊ] is the vowel in *know*, and [f] is the initial sound in *show*. The pairs were designed to be distinctive.

C ₁ iC ₂ i			C ₁ aC ₂ a		
mibi	tibi	sibi	gata	zata	mata
pibi	dibi	kibi	pata	kata	nata

FIGURE 2. Example lexicon in experiment 1.

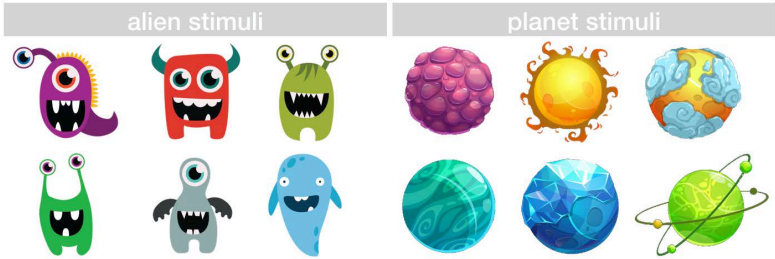


FIGURE 3. Visual stimuli (aliens and planets) in experiment 1.

PROCEDURE. The experiment was implemented on a desktop (adults) or laptop (children) computer using PsychoPy (Peirce 2009). For adult participants, the session took place in a booth in a lab at the University of Edinburgh; the participant was alone in the booth, working through the experiment at their own pace, with all instructions presented in writing. For children, the session took place in a quiet room at the participant’s school, with the experimenter seated beside the child throughout, providing all instructions aloud; the experimenter provided encouragement to the child but no informative feedback. For all participants, the session consisted of two stages: exposure and production. The entire experiment lasted fifteen to twenty-five minutes. In the exposure phase, participants saw ninety-six trials total, twelve presentations each of four aliens and four planets along with their corresponding labels. Trials were either singular or plural. On singular trials (three of twelve for each object), a single object appeared on the screen, and the label for the object was presented auditorily. On plural trials (nine of twelve for each object), a set (either two, four, or six) of the same objects appeared on the screen, and the label and plural marker were presented auditorily. Participants were instructed to pay attention and repeat the description they heard on each trial. Example plural training trials in each condition are shown in Figure 4 and Figure 5.

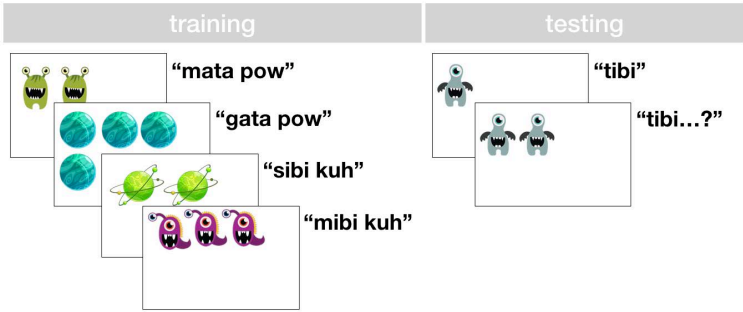


FIGURE 4. Example plural training and testing trials for each class in the phonology-only condition. Testing trial illustrates a new item that matches the phonology of one class. Note that orthographic presentation of words and plural markers is for illustration only.

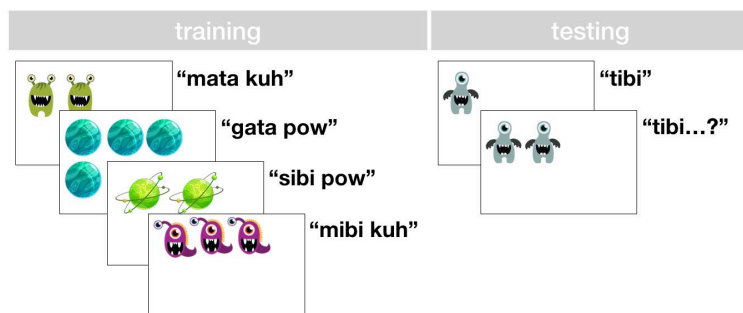


FIGURE 5. Example plural training and testing trials for each class in the semantics-only condition for experiment 1. Testing trial illustrates a new item that matches the semantics of one of the classes.

Note that orthographic presentation of words and plural markers is for illustration only.

In the production phase, participants saw forty-eight trials total, four each of all twelve aliens and planets with their corresponding labels (that is, eight trained items and four new items). On each trial, a singular picture first appeared, and the label for that picture was presented auditorily. Then a plural set of the same objects appeared, and the participant was instructed to provide the description. Unconditional feedback was given in the form of the correct description presented auditorily after the participant had produced their description, thus allowing participants to continue to learn throughout this phase.¹⁰ Example testing sequences for the phonology-only and semantics-only conditions are also illustrated in Fig. 4 and Fig. 5, respectively.

2.2. RESULTS AND DISCUSSION. Figure 6 shows average accuracy in plural-marker production across conditions for familiar (old) items, heard during training, and new items. Data were analyzed using logistic mixed-effects regression models.¹¹ Overall, both adults and children performed above chance in this task for both trial types (as indicated by a significant model intercept; for children, old: $\beta = 0.86 \pm 0.15$, $p < 0.001$; new: $\beta = 0.82 \pm 0.19$, $p < 0.001$; for adults, old: $\beta = 4.76 \pm 0.51$, $p < 0.001$; new: $\beta = 4.12 \pm 0.59$, $p < 0.001$). To compare across age groups and conditions, we ran a model predicting responses by age (adult vs. child), trial type (old vs. new), and condition (phonology-only vs. semantics-only). The model output is shown in Table 1.

This model revealed a significant effect of age and a significant interaction between age and condition. The latter is driven by a small advantage for adults in semantics, which is not present in children.¹² What we take this result to show is that, while adults

¹⁰ There was also a short final judgment phase run in this experiment, in which learners saw a singular picture, heard the description, saw a plural picture, and then had to choose between two descriptions of it. We do not report the results of that phase in the main text here since many children struggled to listen to both descriptions before responding. However, a logistic mixed-effects regression model predicting children's choice based on trial type (old or new item) and condition (phonology-only or semantics-only), both sum coded, revealed a significant effect of trial type ($\beta = -0.19 \pm 0.09$, $p = 0.03$), but no effect of condition ($\beta = -0.33 \pm 0.20$, $p = 0.10$) or their interaction ($\beta = 0.02 \pm 0.09$, $p = 0.84$). The same model predicting adult's choice revealed no significant effects (largest $\beta = -0.46 \pm 0.55$, $p = 0.41$). Both models include random by-participant and by-item (noun and object) intercepts and random slopes for participant by trial type.

¹¹ All models reported throughout were run in R (R Core Team 2017) using the lme4 package (Bates 2010), with by-participant and by-item (object and noun) random intercepts, and a by-participant random slope for trial type when trial type was among the fixed effects. All fixed effects were sum coded.

¹² The significance of this interaction may seem surprising given Fig. 6, but note that the regression analysis is in logit rather than probability space, to account for the fact that the data are binomial. For binomial data, variances are smaller when probability values are near 0 or 1, and consequently differences between

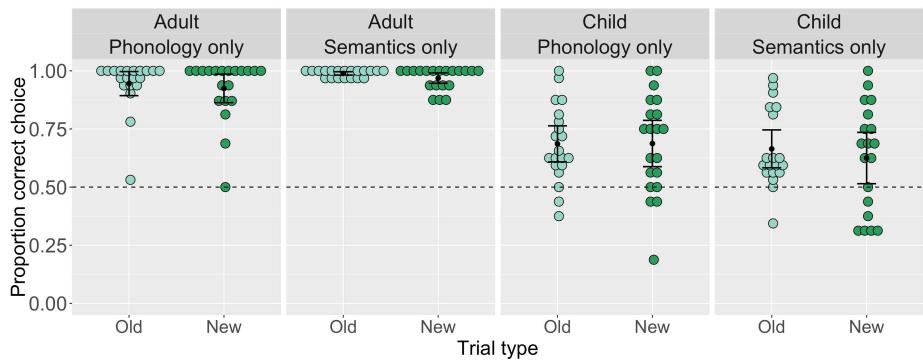


FIGURE 6. Overall accuracy in plural-form production for adults and children (each participant represented by a dot) in both age groups on old (trained) and new (untrained) items in each condition of experiment 1. Error bars represent 95% confidence intervals (CIs) on the mean of by-participant means.

	EST	SE	z-VALUE	p-VALUE	
(intercept)	2.42	0.17	14.17	< 0.001	*
Age	1.54	0.15	10.00	< 0.001	*
Condition	−0.18	0.15	−1.18	0.24	
Trial type	0.16	0.09	1.72	0.09	
Age : Condition	−0.30	0.15	−2.01	0.04	*
Age : Trial type	0.13	0.09	1.50	0.13	
Trial type : Condition	−0.13	0.08	−1.62	0.11	
Age : Condition : Trial type	−0.07	0.08	−0.88	0.38	

TABLE 1. Output of logistic regression model for experiment 1 (* indicates $p < 0.05$).

were uniformly more successful at this task compared to children, the cues are quite closely matched for adults, and they are not learned differently for children. This suggests that children are similarly sensitive to both of these cues in isolation. Figure 7 shows the by-trial averages (collapsing old and new items). These figures indicate that learning was primarily accomplished in the training phrase, with no obvious differences in trajectory across the testing stage in either age group or condition.

3. EXPERIMENT 2. The results of experiment 1 show that, when faced with a novel noun class system in which class is deterministically cued by either animacy or a reduplicative vowel in the noun, children readily learn both. They do not learn quite as successfully as adults, but they are well above chance for both cue types. Moreover, when these cues are independent, adults and children learn both types of cues with similar accuracy.

Having established these two cues as learnable by children in the context of this task, in experiment 2, we follow Culbertson et al. (2017) in asking whether a preference for one cue over the other will emerge when the cues are in conflict. We do this by deliberately confounding the cues during training, such that both animacy and the reduplica-

probability values near 1 or 0 are more meaningful than differences around probability 0.5; running the regression in logit space deals with this asymmetry. The difference between semantics and phonology for adults—0.04 in probability space—translates to a log-odds difference of 1.29. By comparison, that same difference for children (who are not near probability 1) translates to a log-odds difference of just 0.16; in other words, the effect of trial type differs between adults and children in logit space. A separate logistic regression comparing the two conditions in children alone reveals no main effect of condition ($\beta = 0.12 \pm 0.16$, $p = 0.46$). For adults alone there is a marginal effect of condition ($\beta = -0.40 \pm 0.21$, $p = 0.08$), with higher accuracy in the semantic condition. See Agresti 2002 and Jaeger 2008 for a related discussion of why logit regression is preferred to ANOVA for binomial data.

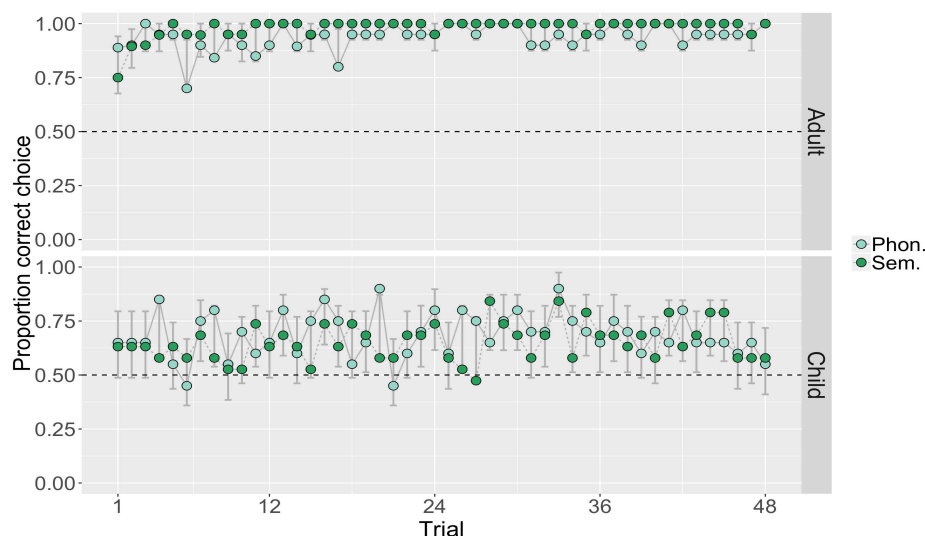


FIGURE 7. Average accuracy across participants in plural-form production for both age groups by test trial in each condition of experiment 1. Error bars represent 95% CIs on the mean of by-participant means.

tive vowel predict noun class equally well, and then test learners on new items where the two cues conflict (i.e. the semantic cue suggests one classification and therefore one marker, but the phonological cue the other marker). By design, the system learners are trained on is therefore ambiguous as to which cue the learner should use when they are in conflict. If the learner has a preference for the phonological cue, then they should use this cue to determine the class of a new item, regardless of its animacy. In contrast, if the learner has a preference for the semantic cue, they should use animacy to determine the class of a new noun and disregard the form of the word itself. Given that we are specifically interested in uncovering whether children show a bias in favor of phonology compared to adults, we also test adult learners under these same conditions. If both adult and child learners are equally likely to go with the semantic or phonological cue, this would suggest that no meaningful bias exists, which is what we might expect for two equal-salience cues based on the results reported in Culbertson et al. 2017. By contrast, if children are significantly more likely to use the phonological cue compared to adults in the same task, this will point to an a priori bias in favor of phonology on the part of child learners.

3.1. METHOD.

PARTICIPANTS. Child participants were twenty six- to seven-year-old children (eight female; mean age = 7;1, range = 6;2–7;11), who were native speakers of English. Children were recruited via the University of Edinburgh's Wee Science Lab Facebook group, or through a local Girl Guide (scouting) troop. All children were local residents in the Edinburgh area. Adult participants were twenty native English-speaking adults (thirteen female), who were students at the University of Edinburgh and received course credit for participation. Three children were bilingual (one each in Gaelic, Portuguese, and Urdu), and six adults were bilingual (four in Hindi/Urdu, one in French and Mandarin, and one in Spanish).¹³

¹³ See n. 8 above regarding noun class systems in these languages. Note here that we generally have fewer bilinguals in this sample.

STIMULI. Visual stimuli, lexical items, and plural markers were identical to those used in experiment 1. As described above, in this experiment, cues to noun class were deliberately confounded, such that class was cued (with equal reliability) by both a semantic and a phonological cue. For example, one class might consist of aliens with labels of the form C₁iC₂i, and the other class the planets with labels of the form C₁aC₂a. The mapping between word-form categories (C₁iC₂i or C₁aC₂a) and animacy categories (aliens, planets) was randomized between participants.

PROCEDURE. Testing took place in a quiet room or a lab booth (for adults), using a laptop computer running PsychoPy (Peirce 2009). The procedure was identical for adults and children, with one exception: as in experiment 1, the experimenter was seated beside child participants throughout and provided instructions and encouragement but no feedback, whereas the experimenter was not in the room with adult participants and instructions were presented on-screen. The details of the experimental procedure differed somewhat from experiment 1, in that the session consisted of three stages (rather than the two in experiment 1): exposure, production with feedback, and production without feedback (an addition to the experiment 1 method). The entire experiment lasted fifteen to twenty-five minutes. In the exposure phase, participants saw ninety-six trials total, twelve each of four aliens and four planets with their corresponding labels. Trials were either singular or plural. On singular trials (three of twelve for each object), a single object appeared on the screen, and the label for the object was presented auditorily. On plural trials (nine of twelve for each object), a set (either two, four, or six) of the same objects appeared on the screen, and the label and plural marker were presented auditorily. Participants were instructed to pay attention and repeat the description they heard on each trial. Example plural training trials are shown in Figure 8.

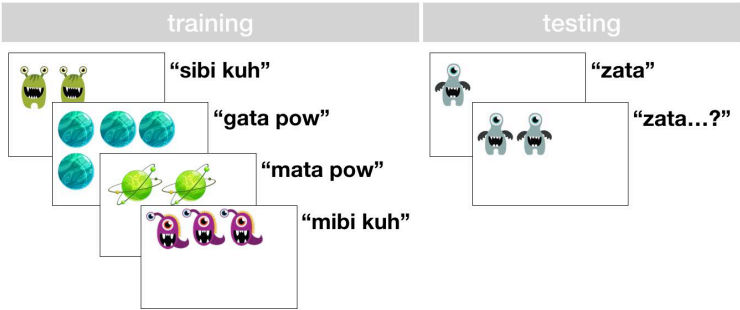


FIGURE 8. Example training and testing trials for each class for experiment 2. In training trials semantic and phonological cues are confounded. Testing trial illustrates a new item with conflicting cues (semantics matches one class, phonology matches the other). Note that orthographic presentation of words and plural markers is for illustration only.

The first production phase gives participants some practice producing descriptions on their own, but also serves as additional training, since feedback was given on each trial. Participants saw twenty-four trials total, featuring the same four aliens and four planets as in the exposure phase, each appearing three times. On each trial, a singular picture first appeared, and the label for that picture was presented auditorily. Then a set of the same objects appeared, and the participant was instructed to provide a description. Unconditional feedback was given in the form of the correct description presented auditorily; thus participants could in principle continue to learn throughout this phase.

In the final phase, participants saw forty-eight trials total, four each of all twelve aliens and planets with their corresponding labels (that is, eight trained items and four

new items). On each trial, a singular picture first appeared, and the label for that picture was presented auditorily. Then a set of the same objects appeared, and the participant was instructed to provide a description. No feedback was provided during this phase.¹⁴ Trained items (i.e. those items encountered earlier in the experiment) have both phonological and semantic cues pointing to one of the two markers deterministically; in other words, for trained items, the cues are *ALIGNED*. Untrained new objects have *CONFLICTING* cues; the phonological cue is used for one class of nouns during exposure, but the semantic cue is used for the other. Responses for conflicting trials can therefore align with either the semantic cue or the phonological cue. An example testing sequence with conflicting cues is shown in Fig. 8 above.

3.2. RESULTS AND DISCUSSION. Below, we focus our analysis on participant responses during the final production phase, in which no feedback is provided. However, the responses provided during the first production phase, in which participants are given unconditional feedback, are also informative. Behavior across these trials can give us a sense for the level of learning participants have reached during exposure, and whether they are learning from the feedback provided here. For both age groups, accuracy is relatively high from the first trials, as shown in Figure 9, suggesting that they have learned the system primarily during exposure (as in experiment 1). Accuracy was generally high on these trials, suggesting that having two aligned cues, which both predict class, leads to better learning of trained items than in the single-cue conditions (cf. Fig. 6). This was confirmed by a logistic regression model comparing familiar (old) items in the production testing phase of experiment 1 with trials in the practice production phase of experiment 2 (recall that these are also all familiar items). The model revealed a significant effect of age ($\beta = 1.20 \pm 0.17$, $p < 0.001$), experiment ($\beta = 0.35 \pm 0.14$, $p = 0.01$), and their interaction ($\beta = -0.46 \pm 0.14$, $p = 0.01$), reflecting the fact that adults benefit less from the presence of two cues than children do.

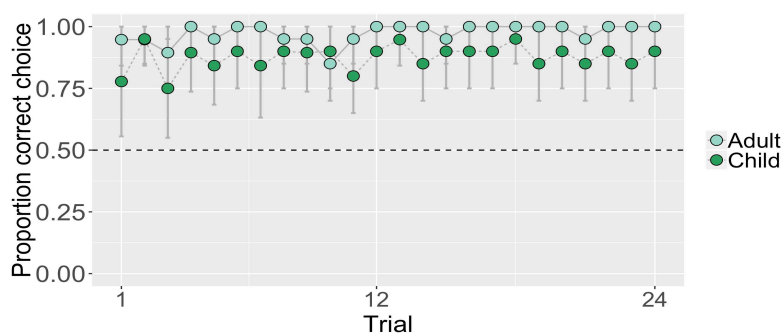


FIGURE 9. Average proportion of correct responses across participants in each age group by trial in the first production stage of experiment 2. Note that in this stage, feedback is provided, and cues are equally predictive. Error bars represent 95% CI on by-participant means.

Figure 10 illustrates the results from the critical second production phase. This figure shows the average proportion of trials on which children and adults chose the semantic cue for aligned and conflicting trial types. Note that unlike in experiment 1, these two trial types measure very different things. Aligned trials are equivalent to old trials above and indicate how accurately participants reproduce the language they were trained on.

¹⁴ Note that the first production phase is similar to the testing phase in experiment 1 (participants produce responses and get feedback). This is shorter, however, and includes only previously seen (aligned) items. The final production phase is the critical test here.

In this case, if a participant's plural marker conformed to the semantic cue, then it also conformed to the phonological cue. Accuracy was generally high on these trials, again suggesting that having two aligned cues, both predicting class, leads to better learning of trained items than in the single-cue conditions (cf. Fig. 6). As in experiment 1, however, there is a difference in performance between adults and children, as indicated by a significant main effect in a logistic mixed-effects model predicting response accuracy by age ($\beta = 2.36 \pm 0.58$, $p < 0.001$).

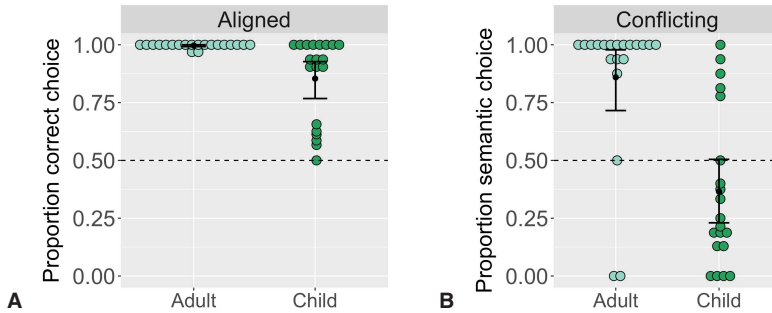


FIGURE 10. A: Average proportion of trials on which adults and children (each participant represented by a dot) choose the correct marker for aligned trials in experiment 2. B: Average proportion of trials on which adults and children (each participant represented by a dot) choose the marker based on the semantic cue for conflicting trials. Error bars represent 95% CIs on the mean of by-participant means.

For conflicting trials, there is no correct answer. Rather, our measure indicates whether responses tended to conform to the semantic or the phonological cue when the two cues were in conflict. In Figure 10B, we follow Culbertson et al. (2017) in arbitrarily choosing to plot the proportion of markers conforming to the semantic cue; a low value therefore indicates use of the marker conforming to the phonological cue. The results in Fig. 10B suggest that adults and children perform differently on conflicting trials. These trials were analyzed using a logistic mixed-effects model predicting semantic choice based on age. The model revealed a significant effect of age ($\beta = 3.19 \pm 0.69$, $p < 0.001$), indicating that children differ from adults in the degree to which they use markers based on the semantic cue. A look at the individual participant data for adults confirms that they were highly likely to base their use of the plural marker on animacy in conflicting trials. Indeed, almost all adult participants did so (seventeen of twenty), many categorically. In contrast, there is substantial individual variation in the children's data, with five children tending to use the semantic cue, one child alternating between the two cues, and the remaining fourteen tending to choose the phonological cue, some very strongly, some less so.

Figure 10A shows that a subset of the children failed to learn the system accurately: they are close to chance on aligned trials. One possibility is that these particular children tend to also produce responses close to chance on conflicting trials, driving the difference found between children and adults. To rule this out, we ran the same analysis but including only participants who performed better than chance on aligned trials (75%, according to an exact binomial test for sixteen trials, leading to the exclusion of six child participants). This analysis replicates the difference across age groups ($\beta = 3.71 \pm 0.92$, $p < 0.001$). Children who learn the system well are likely to consistently choose markers either based on semantics or based on phonology, rather than at chance; the majority of such children use the phonological cue.

4. INTERIM DISCUSSION. Experiment 1 showed that both adults and children are able to learn a novel noun class system based on either a perfectly reliable phonological cue (vowel reduplication) or a perfectly reliable semantic cue (animacy). These cues were in fact learned to a similar level of accuracy within each age group in a task where only one cue is predictive. However, experiment 2 suggests that when these two cues are in conflict, children rely more on phonological cues than adults do. Adults were highly likely to choose the semantic cue in this case, while children were if anything more likely to choose the phonological cue (as indicated by a marginally significant negative model intercept for children's conflicting trials: $\beta = -1.22 \pm 0.68, p = 0.07$).¹⁵ The results of this experiment therefore point to the possibility of an a priori bias on the part of child learners. But there are several alternative interpretations of these data that must be considered.

First, the phonological cue we use (similar to those used in previous studies, e.g. Frigo & McDonald 1998, Culbertson et al. 2017) leads to nouns that are highly similar to one another. If children have more difficulty than adults in telling the similar-sounding nouns apart, this could actually make the task of learning the phonological cue easier; if all nouns in one class sound the same, then children would only have to learn the mapping between a single 'word' and its class marker. By contrast, if the relative similarity of the objects in each class is lower (i.e. the aliens look on average less like each other), then this might take longer to learn, thus leading to a preference for the phonological cue. If children have trouble distinguishing words, then we should see evidence for this in their productions. However, children's average accuracy at reproducing the stems was above 90% (Figure 11A).¹⁶ Further, it is not clear why adults and children would differ in their ability to discriminate our phonological and semantic cues; in other words, it is not clear why the similarity of the nouns would not simply be reflected in a preference for phonology in both age groups.

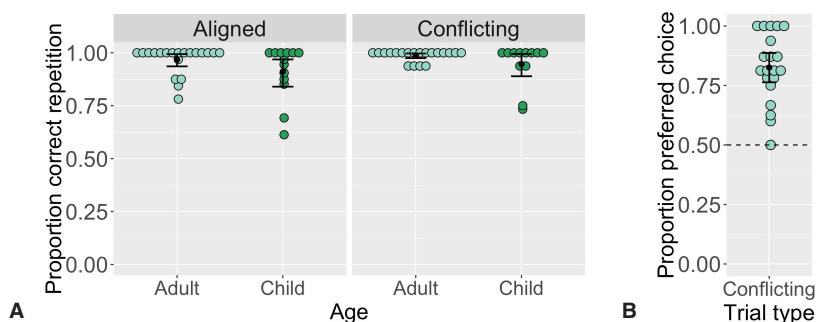


FIGURE 11. A: Average accuracy for adults and children (each participant represented by a dot) at reproducing stems in the production testing phase, by trial type (aligned vs. conflicting) in experiment 2. Recall that in this phase, participants hear the singular, stem-only form, and then must produce that same stem with a plural marker. B: Each child's average proportion use of their preferred cue on conflicting trials (calculated by subtracting scores below 0.50 from 1). Error bars represent 95% CIs on the mean of by-participant means.

¹⁵ Indeed, this is in the same direction as the very small effect we saw in experiment 1, suggesting that the presence of conflicting cues serves to bring out this bias.

¹⁶ Productions were scored by a coder blind to the issue at hand: namely, that children are able to accurately perceive and therefore reproduce the stems.

A referee points out that this is perhaps not a very stringent test of whether children accurately ENCODE these words. Specifically, it could be that producing a noun they have just heard is simple enough, but a test that involves recalling the forms of the nouns would show that children have done a worse job at encoding than adults have.

A second possibility is that children are more uncertain in the face of conflicting cues than adults. While children are marginally more likely to choose the phonological cue over the semantic cue, it is the difference between adults and children that suggests a bias in favor of phonology (or against semantics) in children. However, Fig. 10 also suggests the possibility that children are simply less certain than adults, and therefore more variable in their responses. Adults were strongly bimodally distributed, tending to deterministically choose a marker based either on semantics OR on phonology (this same pattern was found in Culbertson et al. 2017). Children, by contrast, are somewhat less deterministic in their choices (though note this is true mainly for children whose accuracy on aligned trials is low, and the effect of age group is still present when we exclude those children from the analysis). Perhaps in the face of conflicting cues, children are simply choosing at random. If so, then the significant difference between adults and children in experiment 2 would be explained as a consequence of a difference in reaction to conflicting cues, rather than a preference for phonological cues per se. But children do not in general behave randomly on conflicting trials. Figure 11B shows the average proportion of conflicting trials on which children chose the marker based on their preferred cue type—for example, if a child chose based on semantics in 25% of trials, then they used their preferred cue type (i.e. the phonological cue) 75% of the time. Children were well above chance on this measure ($\beta = 1.79 \pm 0.27$, $p < 0.001$), indicating that each child is individually consistent in their responses on conflicting trials, rather than simply responding at random. This suggests that the difference between adults and children in our study is not simply a matter of children's random behavior in the face of conflicting cues.

5. EXPERIMENT 3. Results from experiment 2 suggest that when both semantic and phonological cues to class are present and conflicting, children rely more on phonological cues than adults do. In our final experiment, we ask whether this effect is further exacerbated by the early availability of phonological cues. Culbertson et al. (2017) hypothesized that the explanation for children's overreliance on phonology might come from early access to phonological cues, prior to learning noun meanings. They trained adults on a system with confounded semantic and phonological cues, but where one of those cues was made available to the learner earlier. In order to simulate the hypothesized early availability of phonology cues, an additional initial exposure phase was added in which participants were trained on nouns and class markers alone, without pictured referents; in order to simulate early availability of the semantic cue, in the additional initial phase they trained participants on pictures and markers without the noun itself present. They found that adults tended to rely more on the earlier-available cue in later conflicting trials (see Fig. 1), even if the early-available cue was relatively low salience; importantly, this staging effect boosted both phonological and semantic cues, and was not restricted to the more natural phonology-first staging. Culbertson et al. (2017) therefore argue that the apparent bias toward using (potentially less reliable) phonological cues may reflect developmental changes in the intake, rather than any preference for phonological cues per se (Gagliardi & Lidz 2014).

While the results of experiment 2 suggest that children may in fact rely more on phonological cues than adults do, it could nevertheless be the case that the early availability of phonology also plays a role. In experiment 3, we therefore test whether the effect of cue availability reported for adults in Culbertson et al. 2017 is also found in children. One possibility, in light the findings of experiment 2, is that children may simply prefer to use the phonological cue regardless of its availability. Relatedly, early

availability may further strengthen the phonological cue, but not the semantic cue. However, if cue availability and children's *a priori* preference for phonology are independent of one another, then we expect to see both cues strengthened by our staging manipulation.

5.1. METHOD.

PARTICIPANTS. Participants were forty six- to seven-year-old children (phonology-first: nine female; mean age = 7;1, range = 6;2–8;1; semantics-first: fourteen female; mean age = 7;0, range = 6;0–8;1), who were native speakers of English. Four additional children were run but excluded due to fussiness (three) or technical problems (one). Children were recruited via the University of Edinburgh's Wee Science Lab Facebook group. All children were local residents in the Edinburgh area. Eleven children were bilingual (two in French, Arabic, and Kurdish, two in Dutch, one in Polish, Spanish, and Gaelic, and one each in Finnish, French, Greek, Hungarian, Japanese, and Spanish).¹⁷ Participants were randomly assigned to one of the two conditions.

STIMULI. Visual stimuli, lexical items, and plural markers were identical to those used in experiments 1 and 2. There were two conditions, described in detail below, which differ only in whether there is early exposure to phonology (without semantics) or semantics (without phonology). Following this early exposure, as in experiment 2, cues to noun class were deliberately confounded, such that class was cued (with equal reliability) by both a semantic and a phonological cue. For example, one class might consist of aliens with labels of the form C_1iC_2i , and the other class of the planets with labels of the form C_1aC_2a . The mapping between word-form categories (C_1iC_2i or C_1aC_2a) and animacy categories (aliens, planets) was randomized between participants.

PROCEDURE. Testing took place in a quiet room, using a laptop computer running PsychoPy (Peirce 2009). The experimenter was seated beside the child throughout and provided encouragement but no feedback. The session consisted of four stages: single-cue exposure, full exposure, production with feedback, and production without feedback (i.e. the procedure was identical to experiment 2 apart from the addition of the initial single-cue phase). The entire experiment lasted twenty to twenty-five minutes. In the single-cue exposure phase, there were forty-eight trials total (six each of four aliens and four planets or their corresponding labels), and all trials were plural (i.e. all trials featured the distributional marker of class). In the semantics-first condition, on each trial, a set (either two, four, or six) of the same objects appeared on the screen, and the plural marker was presented auditorily without the preceding noun. Note that this is why all trials are plural in this stage; otherwise in this condition there would be nothing to learn (or repeat) on a singular trial. In the phonology-first condition, on each trial, a label and plural marker were presented auditorily (without any picture). Examples of both types of single-cue exposure are shown in Figure 12. Participants were instructed to pay attention and repeat the description they heard on each trial.

Following single-cue exposure, the remainder of the experiment was identical across conditions. There were forty-eight trials total (six each of four aliens and four planets and their corresponding labels), featuring the same items (either words or pictures) as in the single-cue exposure for that participant. Crucially, this time the second cue was also incorporated, and singular trials were included. On singular trials, a single object appeared

¹⁷ Most of these languages have noun classifications systems of some kind, most featuring a mix of phonological and semantic cues (see n. 8 and n. 13 above). Japanese has a numeral classifier system, based on semantic features alone. Finnish does not have a noun class system.

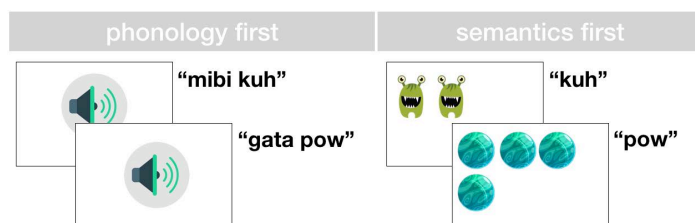


FIGURE 12. Example phonology-first and semantics-first single-cue exposure in experiment 3. Note that orthographic presentation of words and plural markers is for illustration only.

on the screen, and the label for the object was presented auditorily. On plural trials, a set of the same objects appeared on the screen, and the label and plural marker were presented auditorily. Participants were instructed to pay attention and repeat the description they heard on each trial. Note that the total amount of exposure is the same as in both previous experiments (ninety-six trials total across the two exposure stages).¹⁸

Following exposure, participants completed two production phases that were identical to experiment 2: one with feedback (twenty-four trials, three each of the same four aliens and four planets as in the exposure phases), and one without feedback (forty-eight trials total, four each of all twelve aliens and planets with their corresponding labels). Note that as in experiment 2, objects seen during exposure have both the phonological and the semantic cues that point to one of the two markers deterministically. In other words, for trained items, the cues are aligned. The new objects have CONFLICTING cues; the phonological cue used for one class of nouns during exposure, but the semantic cue used for the other. Responses for conflicting trials can therefore align with either the semantic cue or the phonological cue.

5.2. RESULTS AND DISCUSSION. Figure 13 shows children’s response behavior on aligned and conflicting trials in each of the two staged conditions, with the results from (the unstaged) experiment 2 repeated for comparison. As above, we first analyze aligned trials using a logistic mixed-effects model predicting semantic choice based on condition (in this case we use treatment coding, with unstaged as the baseline level, compared to phonology-first and semantics-first). As expected, there was no significant difference between conditions for these trials (phonology-first vs. unstaged: $\beta = -0.42 \pm 0.73$, $p = 0.57$; semantics-first vs. unstaged: $\beta = -0.16 \pm 0.72$, $p = 0.83$); children’s ability to learn the system they were trained on was not affected by staging.

Turning to conflicting trials, where staging is predicted to affect children’s reliance on each cue type, we first compare the two staged conditions. A logistic mixed-effects model predicting semantic choice based on condition (phonology-first, semantics-first) revealed a significant effect of condition ($\beta = -1.55 \pm 0.47$, $p < 0.001$), suggesting that staging indeed has an effect on how likely children are to use the semantic cue. Early exposure to the semantic cue alone leads children to choose markers based on the semantic cue, while early exposure to the phonological cue leads them to choose markers based on phonology when the two cues conflict. Figure 13 suggests that, relative to the unstaged condition, the semantic cue may benefit more from early availability, perhaps because children were already slightly more likely to choose the phonological cue when

¹⁸ This is in contrast with the adult experiments reported in Culbertson et al. 2017, where participants received twice the training in staged experiments. We did this in order to keep the experimental sessions relatively short (around twenty to twenty-five minutes) for our child participants.

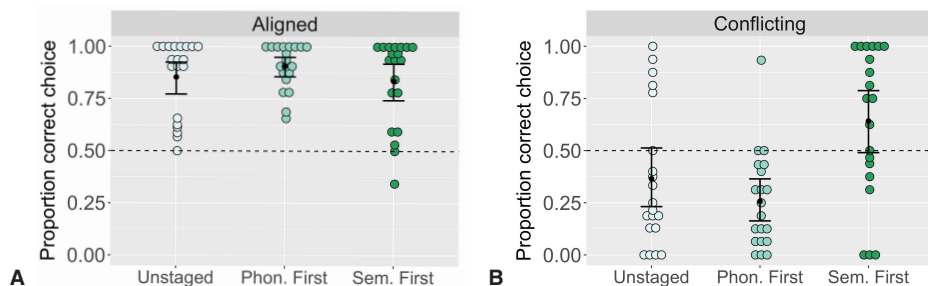


FIGURE 13. A: Average proportion of trials on which children (each represented by a dot) choose the correct marker for aligned trials in unstaged (experiment 2 results, repeated from Fig. 10), phonology-first, and semantics-first conditions (experiment 3). B: Average proportion of trials on which children (each represented by a dot) choose the marker based on the semantic cue for conflicting trials. Error bars represent 95% CIs on the mean of by-participant means.

no staging was present.¹⁹ This is confirmed by a logistic mixed-effects model (again, treatment coding with unstaged as the baseline), which revealed a significant difference between performance on conflicting trials in the unstaged and semantics-first conditions ($\beta = 2.16 \pm 0.87$, $p = 0.01$), but no difference between unstaged and phonology-first ($\beta = -0.82 \pm 0.86$, $p = 0.34$).

6. GENERAL DISCUSSION. The set of experiments reported above aimed to uncover whether child learners have a preference for phonological cues when learning a novel noun class system, and whether they are sensitive to early availability of cues in the same way that the adults in Culbertson et al. 2017 were. In particular, Culbertson et al. (2017) found that adults were sensitive to cue salience: some types of semantic cues (e.g. animacy) were learned more easily than others (e.g. flexibility), and some types of phonological cues were easier than others (e.g. a suffix on the noun vs. both a prefix and suffix together). However, there was no evidence of any bias in adults against phonology; when two cues were both equally predictive in training, but conflicted at test, cue salience rather than cue type appeared to determine which cue adults relied on. Early availability also played a role: a weak phonological cue could be strengthened relative to a strong semantic cue if the phonological cue was available earlier.

Using a similar procedure with child learners, here we have found that two cues—animacy and vowel reduplication—are treated differently by child and adult learners. We chose these two cues because we hypothesized that both were likely to be highly salient. Indeed, when trained on novel noun class systems in which one cue or the other was perfectly predictive, adults and children learned both systems successfully, and to approximately the same level of accuracy. When both cues were equally predictive during training but conflicted at test, however, adults overwhelmingly chose a class marker

¹⁹ Given that experiments 2 and 3 suggest a developmental difference in the reliance on phonology, we also ran a post-hoc analysis investigating whether younger and older children in our sample differ in the extent to which they favor the phonological cue relative to the semantic one. We ran a mixed-effects logistic regression model predicting use of the semantic cue in conflicting trials by condition (unstaged, phonology-first, semantics-first) and children's age (either six or seven years old). In addition to the effects of condition already reported above, this model revealed a marginal effect of age ($\beta = -0.61 \pm 0.34$, $p = 0.07$), with younger children less likely to use the semantic cue across conditions. This is a very small sample size, but further research could therefore extend our age range to investigate whether younger children show an even stronger preference for the phonological cue, while children older than seven resemble adults more closely.

based on the semantic cue, while children were much less likely to do so. Interestingly, while numerically more children chose a marker based on phonology, a subset of older children showed a strong preference for the semantic cue.

We also found that children's classification systems are influenced by early availability. When they were first trained on the semantic cue alone, they were more likely to use that cue later—both compared to unstaged training (both cues available at the same time) and compared to when they were trained on phonology first. When they were trained on the phonological cue first, almost all children used that cue later. But this was not significantly different from the unstaged condition, which already showed a phonology preference. This is in line with the findings of Culbertson et al. (2017). In their study, when cues were of equal salience, staging led adults to choose the early-available cue. When cues were asymmetric—for example, a higher-salience semantic cue and a lower-salience phonological cue—staging strengthened the weaker cue. In this case, we have some evidence that the phonological cue is more readily used than the semantic cue, and the latter benefits more from staging. In the natural language data described above, the phonological cues (e.g. noun endings) are often less reliable than the semantic cues (e.g. natural gender or animacy). Therefore, early availability may benefit phonological cues more in that case.

To summarize, our results suggest that the relative salience of phonological and semantic cues depends on age; relative to adults, children are much more likely to go with a phonological cue if one is present. If phonological cues are also generally available earlier than semantic cues in natural language acquisition, then both of these two factors may contribute to children's crosslinguistic overreliance on phonology. It is worth noting that the early availability of phonological cues is likely to impact not just noun class learning; phonological cues to word classes more generally (e.g. Kelly 1992, Shi et al. 1998) and to morphosyntactic and syntactic dependencies like agreement (e.g. Culbertson et al. 2017) may also be available before syntactic or semantic cues. In other words, the impact of staging on the development of representations is likely widespread. Of course, here we have controlled for cue reliability (both cues are always perfectly reliable) and cue saliency (both cues are highly salient), and it remains for future research to investigate how these factors interact with early cue availability. If early-available phonological cues are demonstrably less reliable than the semantic cues, children will eventually overcome their preference to use the phonological cue (as Gagliardi and Lidz (2014) show for Tsez).

While this study points to two possible mechanisms for the observed crosslinguistic preference for phonological (as opposed to semantic) cues to noun class in children (early availability of phonological cues, and a preference for phonological cues in children), the cause of children's preference for phonological cues remains unknown. Why are children in our study so much more likely to use the phonological cue to determine class compared to adults? Above we discussed and attempted to rule out two explanations that appeal to factors other than phonological bias *per se* to explain the results. First, we discussed the possibility that children perceive the words as having a higher within-class similarity than the objects, leading to easier learning. However, there was no indication from their productions that children had trouble distinguishing the words from one another (although it is still possible that their encoding may be worse). Further, it is not clear why this would differ so dramatically between adults and children. Second, we entertained the possibility that children's reaction to conflicting cues differs qualitatively from adults', leading them to respond essentially at chance when cues conflict. However, we showed that although children's responses were indeed more vari-

able than adults', they are not responding randomly. Rather, their rate of use of their preferred cue type—whether semantic or phonological—was quite systematic. A third possibility, not discussed above, is that there is something special about reduplication for child learners. In other words, children's preference for the phonological cue in experiment 2 might be due to the specific cue we used, rather than to a bias for phonological cues more generally. We cannot rule out this possibility; doing so would of course require us to identify other phonological cues that were similarly learnable in isolation from our semantic cue (e.g. as in experiment 1). Finally, as a referee points out, it could be that because participants in our task were told to repeat the phrases they heard during training, the phonological cue might be better integrated than the semantic cue. If children in general take longer to figure out the system than adults, then perhaps in some sense they are getting phonology first even in experiment 2. This would explain why the effect of staging in the phonology-first condition is not significant. The apparent advantage for semantics in adults (e.g. Fig. 8) would be less straightforward to explain on this account, since adults will also go through a period of time during which they are repeating the phrases aloud but have not yet integrated the semantic cue. The prediction here is that removing the repetition aspect of our task would remove the asymmetry between phonology and semantics—we would be surprised if this was the case but have little basis for that intuition. If, however, the children (and not adults) are still more likely to use the phonology cue independent of vocal repetition, then this implies a genuine advantage for phonology over semantics that requires explanation.

Here we discuss two such potential explanations. First, recent work by Pertsova and Becker (2018) suggests that learners may acquire these two types of cues via different mechanisms. In particular, Pertsova and Becker (2018) find that when adults and children are taught either a semantic cue (animacy) or a phonological cue (number of syllables), both age groups are more likely to be able to verbalize the 'rule' when learning a semantic cue. Further, overall, adults are more likely than children to verbalize the 'rule'.²⁰ They argue that any difference between phonology and semantics may therefore be driven by the fact that adults learn explicitly (in these types of tasks), and the semantic cue is more likely to be explicitly learned. By contrast, children tend to learn implicitly, and the phonological cue is more likely to be learned in this way. We asked adults in experiment 1 to describe the rule ('How did you decide which plural word to use?'). We found that, indeed, thirty-six out of forty adults were able to verbalize the rule, and the four who did not were in the phonology-only condition.²¹ Given that adults were able to learn both cues to a high degree of accuracy, it is perhaps not surprising that they were generally able to verbalize both. However, if children in our task were not in fact learning explicitly and this correlates with cue type, this could potentially explain our results. Of course, this simply restates the original question: why would it be that phonology is more likely to be learned implicitly than semantics?

To answer this, we turn to a second possible explanation for an asymmetry between phonology and semantics, namely, that relationships between elements occurring WITHIN

²⁰ Similarly, Brown et al. (2018) found that adults were more likely than children to be able to learn explicitly in a task involving noun class learning from semantic cues; they also found little evidence of learning once explicit learners had been excluded.

²¹ Children in Pertsova & Becker 2018 were between five and eleven years old, and even in this case, many children were not able to verbalize any rule. Children in our experiments (six years old) were younger on average; we did not ask them to attempt to explain their behavior in any of our experiments. Unfortunately, adults in experiment 2 were run before we were aware of Pertsova and Becker's results, so we did not include any postexperiment question designed to probe awareness.

a given modality are used preferentially by learners, particular children. The class markers and nouns, along with their phonological features, both occur within the auditory modality. By contrast, the objects are presented visually, in a different modality from the markers they are potentially cuing. Another way of putting this is that semantic information (coming from the visual modality in this case) is in some sense external to the local linguistic information present (cf. Perez-Péreira (1991), who refers to phonological cues from nouns as well as distributional cues from markers as INTRALINGUISTIC, and semantic cues as EXTRALINGUISTIC information). It may be that, for children, the presence of a local, within-domain dependency (i.e. the phonological cue) blocks the formation of non-local, across-domain dependencies. Under this explanation, children are able to acquire either cue type when it is the only cue available, but if a phonological cue is present and suffices to predict the marker, then children will not readily override this cue with one coming from the visual domain. Adults, by contrast, will readily do so. This preference for local, domain-internal cues would also explain the relatively small effect of staging in the phonology-first condition: local phonological cues are available earlier by default.

Research on cue integration across domains may provide some support for the idea that external cues readily used by adults are not used by children to the same degree. For example, Thompson and Massaro (1994) investigated four- and nine-year-old children's use of conflicting speech and gesture (pointing) cues in an ambiguous referential comprehension task. They found that for both age groups, speech had a far greater influence than gesture, though gesture was used more by the older children. Snedeker and Trueswell (2004) showed that, in order to resolve temporary ambiguities in a sentence, five-year-old children use lexical cues (i.e. information about likely arguments of a verb), but fail to use contextual information (i.e. whether there is a need to distinguish between entities present). These studies suggest that when there is conflict, children may preferentially use lower-level cues (i.e. speech rather than gesture, and syntactic rather than contextual information). Stronger confirming evidence of the connection with cue locality would come from a task showing, for example, that children attend to phonological evidence (e.g. the initial syllable of a word) but ignore evidence about the likely semantic properties of an object in on-line comprehension.²²

Interestingly, a preference for local, domain-internal cues may also help to explain why children appear to use noun-external distributional cues like determiners or other agreeing modifiers less than would be expected (given that they are, in fact, perfectly reliable cues to class). For example, Karmiloff-Smith (1981) reports that French children under five years old use phonological information (e.g. a suffix) rather than a previously supplied indefinite article explicitly indicating an object's grammatical gender (e.g. *un* or *une*) to determine which definite article to use with that object (e.g. *le* or *la*). This leaves open the question of whether children would rely on noun-external distributional cues (e.g. the class markers themselves) rather than noun semantics. This is perhaps supported by the fact that research on acquisition of classifier systems (where phonological cues are not present) consistently finds that children tend to overuse general classifiers (e.g. the default-like *go* in Cantonese) before acquiring the semantic

²² For example, using the visual-world paradigm, one could present a story about a girl who really likes to eat sweets, and then present participants with an utterance like 'The girl decided that next she would eat a /kær/ ...'. If child participants show more looks to an image of a carrot relative to an image of a caramel (both of which are phonologically compatible with /kær.../, and both of which are possible arguments of *eat*, but only one of which is consistent with the context that the girl likes sweets) compared to adults, then this would suggest that they are failing to integrate the contextual semantic cue to the same degree that adults would. Of course, the words would need to be frequency-matched, which *carrot* and *caramel* are not.

cues necessary for adult-like production (e.g. Gandour et al. 1984, Yamamoto & Keil 2000, Tse et al. 2007).

These explanations, of course, rest on the generalizability of the results we have presented here. We have used two particular examples of these cue types, both highly salient. Other such matched cues would, by hypothesis, lead to results similar to those we have reported here. However, it is likely that a sufficiently subtle phonological cue would lead children to use a high-salience semantic cue more. In other words, relative salience of the cues is still likely to influence children's ease of acquisition of noun class systems. Further, we have not manipulated cue reliability at all here, nor have we investigated whether the presence of cues that are relevant only for a subset of nouns (e.g. natural gender) impacts learning. There is no doubt that these aspects of the system will play an important role in how noun classification systems are acquired.

7. CONCLUSION. This article started from a well-documented finding in natural language acquisition: children across a number of languages have shown an apparent preference for using phonological cues to determine the class of novel nouns. In previous work, Culbertson et al. (2017) argued that the underlying explanation for this pattern of behavior was likely to be the early availability of noun-internal phonological cues. Such phonological cues can be accessed before any noun meanings are acquired, and therefore children's early representations of noun class may be based on them, with semantic cues only fully integrated later. Using artificial language learning experiments, Culbertson et al. (2017) showed that adult learners attended to both types of cues if both were present simultaneously. However, as predicted, they preferentially used an earlier-acquired cue of either type. This suggested that the apparent preference for phonology in children may not reflect any special bias in favor of phonology per se. Here, we extended these findings to child learners. We showed that, as with adults, early-available cues are preferentially used in later learning. However, we also found evidence for a bias to use phonological cues in children. While this result is somewhat surprising, it raises the interesting possibility that an a priori bias for phonology is present in children. We sketched a potential explanation for why children might favor these cues, based on the fact that they are essentially local; both the trigger (i.e. noun phonology) and the result of class (i.e. agreement on determiners and other elements in the utterance) are present in the speech stream. Semantic cues require the child to integrate external information. This explanation remains to be validated, and we have presented a number of alternative possibilities above.

Children's apparent overreliance on phonological cues in noun class learning has been something of a mystery since Karmiloff-Smith's (1981) seminal work. This is partly due to the intuition that the relevant semantic features are systematically both highly reliable and perceptually (or cognitively) salient. Evidence that some combination of their early availability and an a priori bias may give phonological cues primacy gives us a possible explanation for this pattern of behavior in early acquisition. However, it leaves open the question of why semantic cues nevertheless remain ubiquitous in all noun classification systems found in the world's languages. Put simply, if phonological cues dominate in early learning, then why are semantic cues not obviously eroded over time? One possibility is that systems that mix semantic and phonological cues emerged from purely semantic precursors (Aikhenvald 2000) and now retain only the most learnable (salient, reliable, etc.) vestiges of semantics. Another possibility is that adult users, for whom semantics is clearly a strong cue, play a role in maintaining this aspect of the system they have learned. Careful quantitative work on the history of

noun classification systems, combined with experimental tests of these hypotheses, offer a clear way forward in exploring the complex causal chain of emergence, transmission, and change in this domain.

REFERENCES

- AGRESTI, ALAN. 2002. *Categorical data analysis*. 2nd edn. Hoboken, NJ: John Wiley & Sons.
- AIKHENVALD, ALEXANDRA Y. 2000. *Classifiers: A typology of noun categorization devices*. New York: Oxford University Press.
- ARNON, INBAL, and MICHAEL RAMSCAR. 2012. Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* 122.292–305. DOI: 10.1016/j.cognition.2011.10.009.
- ASHER, R. E. 1985. *Tamil*. (Croom Helm descriptive grammars.) London: Croom Helm.
- AYOUN, DALILA. 2018. Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies* 28.113–48. DOI: 10.1017/S095926951700014X.
- BATES, DOUGLAS. 2010. lme4: Mixed-effects modeling with R. Online: <http://lme4.r-forge.r-project.org/book/>.
- BECKER, MISHA. 2009. The role of NP animacy and expletives in verb learning. *Language Acquisition* 16.283–96. DOI: 10.1080/10489220903178997.
- BERKO, JEAN. 1958. *The child's learning of English morphology*. Cambridge, MA: Radcliffe College dissertation.
- BRAINE, MARTIN D. S. 1987. What is learned in acquiring word classes: A step toward an acquisition theory. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 65–87. Hillsdale, NJ: Lawrence Erlbaum.
- BROWN, HELEN; KENNY SMITH; ANNA SAMARA; and ELIZABETH WONNACOTT. 2018. Semantic cues in language learning: An artificial language study with adult and child learners. PsyArXiv preprint. DOI: 10.31234/osf.io/7hq2c.
- BUNGER, ANN, and JEFFREY LIDZ. 2006. Constrained flexibility in the acquisition of causative verbs. *Proceedings of the Boston University Conference on Language Development (BUCLD)* 30.60–71.
- CARROLL, SUSANNE E. 1999. Input and SLA: Adults' sensitivity to different sorts of cues to French gender. *Language Learning* 49.37–92. DOI: 10.1111/1467-9922.00070.
- COMRIE, BERNARD. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- CORBETT, GREVILLE G. 1991. *Gender*. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 2013. Systems of gender assignment. *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath, Ch. 32. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: <https://wals.info/chapter/32>.
- CULBERTSON, JENNIFER; ANNIE GAGLIARDI; and KENNY SMITH. 2017. Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92.343–58. DOI: 10.1016/j.jml.2016.08.001.
- CULBERTSON, JENNIFER, and COLIN WILSON. 2013. Artificial grammar learning of shape-based noun classification. *Proceedings of the 35th annual meeting of the Cognitive Science Society (CogSci 2013)*, 2118–23. Online: <https://mindmodeling.org/cogsci2013/papers/0389/paper0389.pdf>.
- DEMUTH, KATHERINE. 2000. Bantu noun class systems: Loanword and acquisition evidence of semantic productivity. *Systems of nominal classification*, ed. by Gunter Senft, 270–92. Cambridge: Cambridge University Press.
- DEMUTH, KATHERINE, and DAVID ELLIS. 2008. Revisiting the acquisition of Sesotho noun class prefixes. *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, ed. by Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura, and Şeyda Özçalışkan, 93–104. New York: Psychology Press.
- DENNY, J. PETER. 1976. What are noun classifiers good for? *Chicago Linguistic Society* 12. 122–32.
- DIXON, R. M. W. 1986. Noun classes and noun classification in typological perspective. *Noun classes and categorization*, ed. by Colette G. Craig, 105–12. Philadelphia: John Benjamins.

- FRIGO, LENORE, and JANET L. McDONALD. 1998. Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language* 39. 218–45. DOI: 10.1006/jmla.1998.2569.
- GAGLIARDI, ANNIE. 2012. *Input and intake in language acquisition*. College Park: University of Maryland dissertation.
- GAGLIARDI, ANNIE; NAOMI H. FELDMAN; and JEFFREY LIDZ. 2017. Modeling statistical insensitivity: Sources of suboptimal behavior. *Cognitive Science* 41.188–217. DOI: 10.1111/cogs.12373.
- GAGLIARDI, ANNIE, and JEFFREY LIDZ. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90.58–89. DOI: 10.1353/lan.2014.0013.
- GANDOUR, JACK; SORANEE HOLASUIT PETTY; ROCHANA DARDARANANDA; SUMALEE DE-CHONGKIT; and SUNEK MUKNGOEN. 1984. The acquisition of numeral classifiers in Thai. *Linguistics* 22(4).455–80. DOI: 10.1515/ling.1984.22.4.455.
- HENZL, V. M. 1975. Acquisition of grammatical gender in Czech. *Papers and Reports on Child Language Development* 10.188–98.
- JAEGER, T. FLORIAN. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.434–46. DOI: 10.1016/j.jml.2007.11.007.
- KARMILOFF-SMITH, ANNETTE. 1981. *A functional approach to child language: A study of determiners and reference*. New York: Cambridge University Press.
- KELLY, MICHAEL H. 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99.349–64. DOI: 10.1037/0033-295X.99.2.349.
- KEMPE, VERA, and PATRICIA J. BROOKS. 2001. The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning* 51.221–56. DOI: 10.1111/1467-9922.00154.
- LAKOFF, GEORGE. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- LEVY, YONATA. 1983. It's frogs all the way down. *Cognition* 15.75–93. DOI: 10.1016/0010-0277(83)90034-3.
- LEVY, YONATA. 1988. On the early learning of formal grammatical systems: Evidence from studies of the acquisition of gender and countability. *Journal of Child Language* 15. 179–87. DOI: 10.1017/S0305000900012125.
- LYSTER, ROY. 2006. Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies* 16.69–92. DOI: 10.1017/S0959269506002304.
- MACWHINNEY, BRIAN. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43.1–123. DOI: 10.2307/1166047.
- MARISCAL, SONIA. 2009. Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language* 36.143–71. DOI: 10.1017/S0305000908008908.
- MASICA, COLIN P. 1993. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- MILLS, ANNE E. 1985. The acquisition of German. *The crosslinguistic study of language acquisition*, vol. 1, ed. by Dan Isaac Slobin, 141–254. Hillsdale, NJ: Lawrence Erlbaum.
- OTA, MITSUHIKO, and BARBORA SKARABELA. 2016. Reduplicated words are easier to learn. *Language Learning and Development* 12.380–97. DOI: 10.1080/15475441.2016.1165100.
- OTA, MITSUHIKO, and BARBORA SKARABELA. 2017. Reduplication facilitates early word segmentation. *Journal of Child Language* 45.204–18. DOI: 10.1017/S0305000916000660.
- PEIRCE, JONATHAN W. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2:10. DOI: 10.3389/neuro.11.010.2008.
- PÉREZ-PEREIRA, MIGUEL. 1991. The acquisition of gender: What Spanish children tell us. *Journal of Child Language* 18.571–90. DOI: 10.1017/S0305000900011259.
- PERTSOVA, KATYA, and MISHA BECKER. 2018. Is there a phonological bias in implicit learning of allomorphy? Paper presented at the 92nd annual meeting of the Linguistic Society of America, Salt Lake City, January 2018.
- PLASTER, KEITH; MARIA POLINSKY; and BORIS HARIZANOV. 2013. Noun classes grow on trees: Noun classification in the North-East Caucasus. *Language typology and historical*

- contingency: In honor of Johanna Nichols*, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, and Alan Timberlake, 153–70. Amsterdam: John Benjamins.
- POLINSKY, MARIA, and DAN JACKSON. 1999. Noun classes: Language change and learning. *Cognition and function in language*, ed. by Barbara Fox, Dan Jurafsky, and Laura Michaelis, 29–50. Stanford, CA: CSLI Publications.
- R CORE TEAM. 2017. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: <http://www.r-project.org/>.
- RODINA, YULIA, and MARIT WESTERGAARD. 2012. A cue-based approach to the acquisition of grammatical gender in Russian. *Journal of Child Language* 39.1077–1106. DOI: 10.1017/S0305000911000419.
- SCHULER, KATHRYN D.; CHARLES YANG; and ELISSA L. NEWPORT. 2016. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. *Proceedings of the 38th annual meeting of the Cognitive Science Society (CogSci 2016)*, 2321–26. Online: <https://mindmodeling.org/cogsci2016/papers/0402/paper0402.pdf>.
- SCHWAB, JESSICA F.; CASEY LEW-WILLIAMS; and ADELE E. GOLDBERG. 2018. When regularization gets it wrong: Children over-simplify language input only in production. *Journal of Child Language* 45.1054–72. DOI: 10.1017/S0305000918000041.
- SENFT, GUNTER. 2000. *Systems of nominal classification*. Cambridge: Cambridge University Press.
- SHI, RUSHEN; JAMES L. MORGAN; and PAUL ALLOPENNA. 1998. Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language* 25.169–201.
- SNEDEKER, JESSE, and JOHN C. TRUESWELL. 2004. The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology* 49.238–99. DOI: 10.1016/j.cogpsych.2004.03.001.
- STRICKLAND, BRENT. 2017. Language reflects ‘core’ cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science* 41.70–101. DOI: 10.1111/cogs.12332.
- TARABAN, ROMAN. 2004. Drawing learners’ attention to syntactic context aids gender-like category induction. *Journal of Memory and Language* 51.202–16. DOI: 10.1016/j.jml.2004.03.005.
- THOMPSON, LAURA A., and DOMINIC W. MASSARO. 1994. Children’s integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology* 57.327–54. DOI: 10.1006/jecp.1994.1016.
- TSANG, CARA, and CRAIG G. CHAMBERS. 2011. Appearances aren’t everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.1065–80. DOI: 10.1037/a0023601.
- TSE, SHEK KAM; HUI LI; and SHING ON LEUNG. 2007. The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language* 34.495–517. DOI: 10.1017/S0305000906007975.
- YAMAMOTO, KASUMI, and FRANK KEIL. 2000. The acquisition of Japanese numeral classifiers: Linkage between grammatical forms and conceptual categories. *Journal of East Asian Linguistics* 9.379–409. DOI: 10.1023/A:1008308724059.

Culbertson
Dugald Stewart Building
3 Charles Street
Edinburgh EH8 9AD, UK
[jennifer.culbertson@ed.ac.uk]
[Hanna.Jarvinen@ed.ac.uk]
[frances.i.haggarty@gmail.com]
[kenny.smith@ed.ac.uk]

[Received 10 May 2018;
revision invited 5 August 2018;
revision received 1 October 2018;
accepted pending revisions 11 December 2018;
revision received 15 December 2018;
accepted 17 December 2018]